A major purpose of the Technical Information Center is to provide the broadest dissemination possible of information contained in DOE's Research and Development Reports to business, industry, the academic community, and federal, state and local governments.

Although a small portion of this report is not reproducible, it is being made available to expedite the availability of information on the research discussed herein.

1

LA-UR--84-1499

DE84 012626

NOTICE

PORTIONS OF THIS REPORT ARE ILLEGIBLE. It has been reproduced from the best available copy to permit the broadest possible availability.

Los Alamos National Laboratory is operated by the University of California for the United States Department of Energy under contract W-7405-ENG-36

TITLE INVESTIGATION OF AN EMPIRICAL PROBABILITY MEASURE BASED TEST FOR MULTIVARIATE NORMALITY

AUTHOR(S) J. M Booker, M. E. Johnson and R. J. Beckman

SUBMITTED TO To be presented at the American Statistical Association Annual meeting in Philadelphia, PA, August 13-17, 1984

Los Alamos Los Alamos National Laboratory Los Alamos, New Mexico 87545

FORM NO. 836 R4

INVESTIGATION OF AN EMPIRICAL PROBABILITY MEASURE BASED TEST FOR MULTIVARIATE NORMALITY

J. M. Booker, M. E. Johnson and R. J. Beckman

ABSTRACT

Foutz (1980) derived a goodness of fit test for a hypothesis specifying a continuous, p-variate distribution. The test statistic is both distribution-free and independent of p. In adapting the Foutz test for multivariate normality, we consider using χ^2 and rescaled beta variates in constructing statistically equivalent blocks. The Foutz test is compared to other multivariate normality tests developed by Hawkins (1981) and Malkovich and Afifi (1973). The set of alternative distributions tested include Pearson type II and type VII, Johnson translations, Plackett, and distributions arising from Khintchine's theorem. Univariate alternatives from the general class developed by Johnson et al. (1980) were also used. An empirical study confirms the independence of the test statistic on p even when parameters are estimated. In general, the Foutz test is less conservative under the null hypothesis but has poorer power under most alternatives than the other tests.

I. INTRODUCTION

Foutz (1980) developed a test for goodness of fit for the hypothesis specifying a continuous p-variate distribution. The test statistic, F_n, is based on the supremum of the absolute differences between the hypothesized probability distribution, $\hat{P}_H$, and the empirical probability distribution, $\hat{P}_n$. In contrast to Kolmogorov-Smirnov and Cramer-von Mises empirical distribution function (EDF) tests which compare hypothesized and empirical probabilities of only events in a restricted Borel set, the Foutz test compares hypothesized and empirical probabilities of all possible events.

In this paper, a simulation study is presented to assess the performance of Foutz' test in practical situations including small sample sizes, variety of alternative distributions and most importantly, in the case that the parameters of the hypothesized distribution are estimated. A previous study by Franke and Jayachandran (1983) investigated the power of Foutz' test compared with the chi-square and Kolomogorov-Smirnov tests for normality when the samples were generated from various univariate distributions.

II. F_n AND STATISTICIALLY EQUIVALENT BLOCKS

The F_n statistic is defined by

$$F_n = \sup_{B \in \mathcal{B}(R^p)} |\hat{P}_H(B) - \hat{P}_n(B)|, \tag{1}$$

where $\hat{P}_H(B)$ is the hypothesized probability measure of event B, $\hat{P}_n(B)$ is the empirical probability measure based on the sample, and $\mathcal{B}(R^p)$ denotes the collection of Borel sets in R^p. Foutz (1980) showed that the null distribution of F_n is independent of $\hat{P}_H$ and p.

In order for the empirical probability distribution to be a continuous distribution, and thus, to be comparable to the continuous hypothesized distribution, the total mass from the n discrete data points is read over 'statistically equivalent blocks' defined from the sample. According to Anderson (1966), the sample is used to partition the Euclidean p-space, R^p, by means of cutting functions. In the univariate case n such blocks can be formed from the order statistics of a sample of size n-1 as $B_1=(-\infty,X_{(1)})$, $B_2=(X_{(1)},X_{(2)})$,

..., $B_n=(X_{(n-1)},\infty)$. In the multivariate normal case, n blocks can be obtained from the ordered quadratic forms vis a vis $B_1=(-\infty,Q_{(1)})$, $B_2=(Q_{(1)},Q_{(2)})$, ..., $B_n=(Q_{(n-1)},\infty)$ where

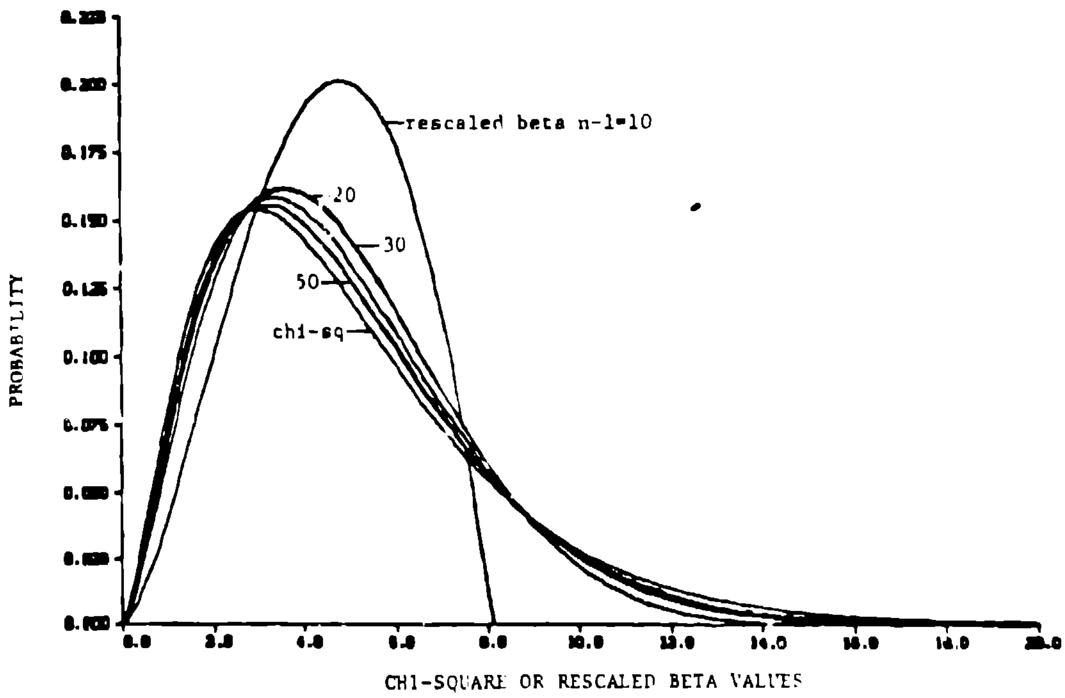
$$Q_{(i)}=(X_{(i)}-\mu)'\Sigma^{-1}(X_{(i)}-\mu). \tag{2}$$

The probability measure for these blocks comes from the chi-square distribution with n-2 degrees of freedom if the sample is from the multivariate normal and the parameters μ and Σ are known or specified. If estimates for the means and covariance matrix are used, then the distribution of $Q_{(i)}$ is only asymptotically chi-square. For samples with n-1 less than p/2, the chi-square distribution holds. However, for large p cases, the quadratic form, $Q_{(i)}$, follows a rescaled beta distribution:

$$\frac{[(n-1)-1]^2}{n-1}\text{ beta }\left(\frac{p}{2},\frac{(n-1)-p-1}{2}\right). \tag{3}$$

Figure 1 shows the differences in the chi-square and rescaled beta distributions for various sample sizes, n-1, for p=5 and p=30 respectively. Using the chi-square distribution when the rescaled beta should be used results in an abnormally high percent rejection of the null hypothesis. This is true of all the multivariate normal tests investigated. The tests in this paper used the chi-square approximation only where appropriate, where sample sizes were large relative to p. Otherwise, the rescaled betas were used.

P=5



P = 30

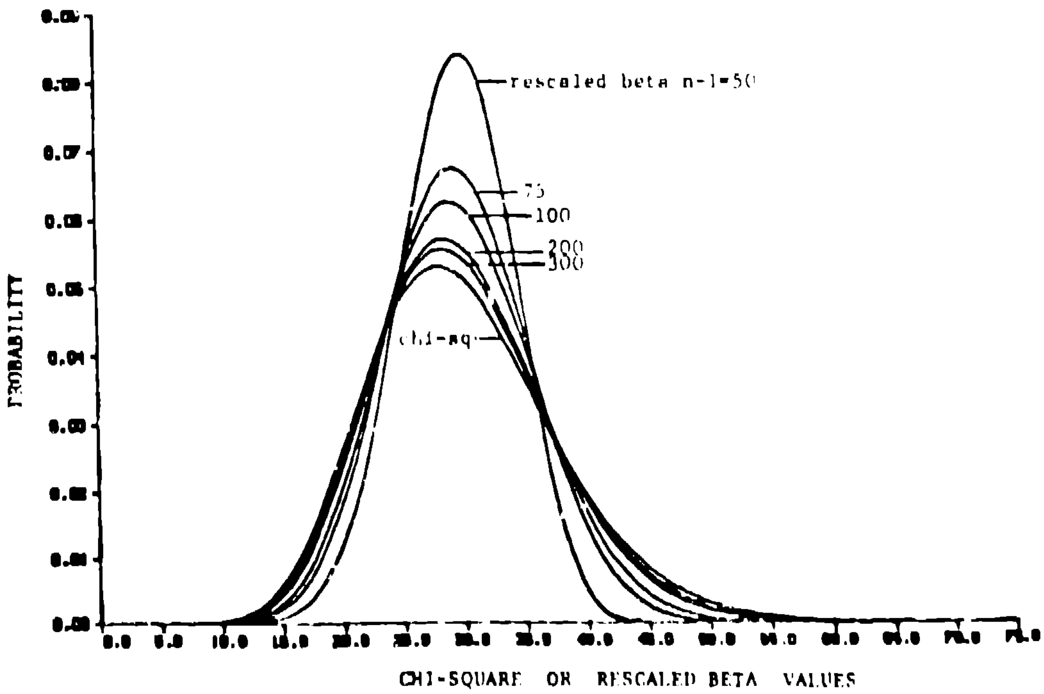


Figure 1

III. EMPIRICAL STUDIES OF F_n

Foutz derived the asymptotic null distribution for F_n as normal with a mean of e^{-1} and variance of $(2e^{-1}-5e^{-1})/(n-1)$.

The test size of F_n was investigated for a variety of n and p combinations. The purpose of this exercise was to test the independence of F_n on p and to gauge the asymptotic behavior in small samples. In this empirical study, the mean and variance parameters were estimated. These estimates were used in the formation of the equivalent blocks.

Table I shows the test size (percent rejection of the null hypothesis) as a function of p for 1000 samples of p-variate normals using various sample sizes. The columns labeled 1%, 5% and 10% refer to the 1%, 5% and 10% asymptotic critical values derived by Foutz which were used to determine the rejection percentages. For convenience, the generated variates are p independent N(0,1). However, by virtue of the calculation of F_n using maximum likelihood estimators for $\underline{\mu}$ and Σ, the reported results hold for arbitarily specified population values of $\underline{\mu}$ and Σ. The test size is equal to or more conservative than the α-level chosen. There appears to be no dependence on n-1 or on p. Therefore, even for relatively small samples, the Foutz test statistic follows its asymptotic normal distribution reasonably well.

TABLE I

PERCENT REJECTION OF THE NULL HYPOTHESIS FOR N-1 AND P COMBINATIONS

USING FOUTZ' ASYMPTOTIC CRITICAL VALUES

N-1	P	10%	5%	1%
50	2	6.70	3.70	0.90
50	10	5.60	2.70	0.90
50	20	8.10	4.70	0.80
50	30	7.20	3.20	1.20
50	40	7.20	3.50	0.70
50	50	8.30	4.30	0.90
20	2	3.80	1.90	0.20
40	2	6.50	3.30	0.60
100	2	9.00	4.60	1.30
200	2	8.30	3.90	0.90
500	2	9.60	5.30	1.10
750	2	9.50	4.40	0.90
100	20	9.10	4.10	0.60
500	20	8.40	4.70	0.80
750	50	11.50	4.80	0.80

Malkovich and Afifi (1973) derived a union-intersection test for multivariate normality. They recommend that their test statistic, denoted here by MAKS, be compared to the Kolomogorov-Smirnov (K-S) critical values. We have found, however, that this leads to an extremely conservative test under the null hypothesis. This observation necessitated the construction of table II which provides alternative critical values for the power studies in Section V.

A third test used for comparison with F_n and MAKS is the Hawkins (1981) test. This test is based on the Anderson-Darling statistic and detects deviations from normality as well as heteroscedastic normal cases.

Another issue related to the MAKS test is the use of the rescaled beta probabilities for the quadratic form instead of the usual chi-square distribution. Table III illustrates the dramatic increase in percent rejection for all three tests when the chi-square is erroneously used. Thus we generally use the rescaled beta in our calculations.

IV. UNIVARIATE CASE STUDIES

Foutz (1980) tested the power of F_n and the chi-square and Kolomogorov- Smirnov tests for cases of a mixed uniform distribution. Using samples of size 50, he concludes that F_{51} performs better than the other two tests in 'some situations' against a normal null hypothesis.

TABLE II

MAKS CRITICAL VALUES GENERATED FROM

10,000 REPLICATIONS OF P-VARIATE NORMALS OF SAMPLE SIZE 50

P	90th Percentile	95th Percentile	99th Percentile
1	0.1374	0.1535	0.1839
2	0.1274	0.1414	0.1692
3	0.1236	0.1366	0.1632
4	0.1202	0.1316	0.1562
5	0.1187	0.1308	0.1549
7	0.1179	0.1290	0.1515
8	0.1183	0.1298	0.1504
10	0.1206	0.1319	0.1518
12	0.1238	0.1343	0.1550
15	0.1322	0.1419	0.1636
20	0.1493	0.1596	0.1809
25	0.1715	0.1822	0.2003

TABLE III

PERCENT REJECTION FOR 1000 20-VARIATE NORMAL SAMPLES OF SIZE 50

COMPARISON OF RESCALED BETA AND CHI-SQUARE DISTRIBUTIONS

Test	10%	5%	1%
Rescaled Beta			
Foutz	8.10	4.70	0.80
Hawkins	0.10	0.00	0.00
MAKS	0.90	0.00	0.00
Chi-square			
Foutz	14.50	8.20	1.00
Hawkins	49.80	30.70	11.10
MAKS	27.90	14.50	2.30

Franke and Jayachandran (1983) run the same three tests on samples from the family of asymmetric stable distributions, mixtures of normals, and the Pearson type I and II distributions. They conclude that F_n does better when 'the parent distribution is heavy-tailed'. However, the chi-square test does better in most of the mixture of normals cases and the asymmetric stable distributions.

Considering the above conflicting results, a general class of distributions developed by Johnson et al. (1980) for simulation studies was used in the following power study. This distribution is unimodal, and symmetric about location parameter, μ, and has the following density:

$$f(x)=\frac{A}{2\sigma\Gamma(\alpha)}\int_{[(A/\sigma)|x-\mu|]^{1/\tau}}^{\infty} w^{\alpha-\tau-1}\exp(-w)dw, \tag{4}$$

$-\infty<x,\ \mu<\infty,\ \alpha,\tau,\sigma>0$, where $A=[\Gamma(\alpha+2\tau)/3\Gamma(\alpha)]^{1/2}$, μ and σ are location and scale parameters respectively, and α and τ are shape parameters. Many special cases emerge for certain α, τ combinations. The normal has $\alpha=1.5$, $\tau=.5$; the uniform has $\alpha=1.0$, $\tau\to0$; the Laplace or double exponential has $\alpha=2$, $\tau=1$.

Figure 2 shows four plots of percent rejection of the null hypothesis using 5% critical values for the Foutz, MAKS, and Hawkins tests versus τ. Each plot has a different value of α. Note that the above special distributions are contained within these four plots. In all cases, the Foutz test has much lower power than the other two tests. In the null case, the MAKS and Hawkins tests are more conservative than F_n.

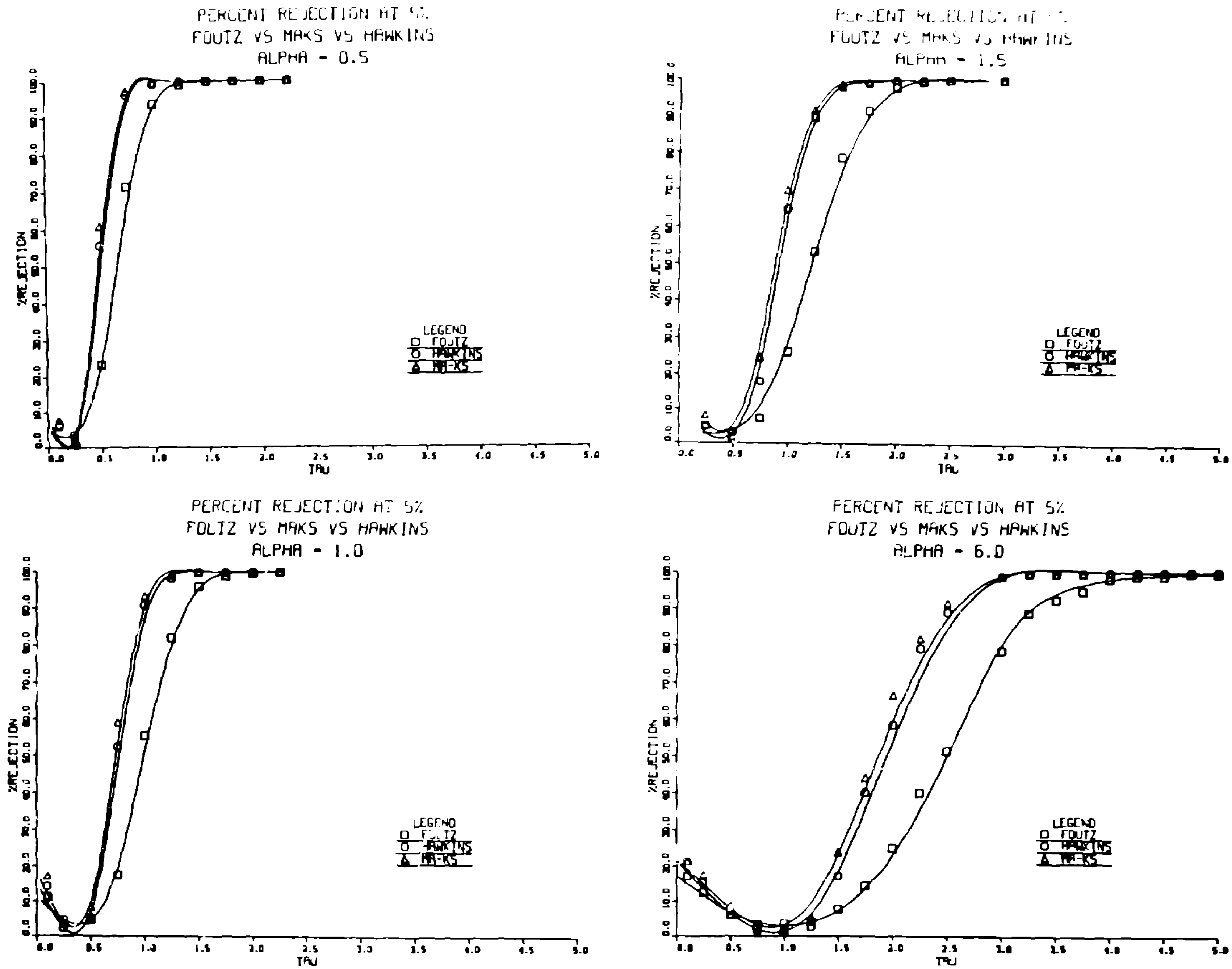

PERCENT REJECTION AT 5%
FOUTZ VS MAKS VS HAWKINS
ALPHA - 0.5
ALPHA - 1.5
ALPHA - 1.0
ALPHA - 6.0
%REJECTION
TAU
LEGEND
FOUTZ
HAWKINS
MA-KS

Figure 2

The results in figure 2 combined with those of previous studies indicate the relatively poor power of F_n in the univariate case against various alternatives.

V. MULTIVARIATE STUDY OF F_n

The following bivariate distributions were used to generate the 1000 samples of size 50 for testing the power of F_n, MAKS and Hawkins against the normal null hypothesis:

- bivariate normal
- Pearson type II
- Pearson type VII
- distribution based on Khintchine's theorem
- Johnson translation system -- lognormal and $\sinh^{-1}$
- Plackett with normal or uniform marginals
- Morgenstern's with normal or uniform marginals

The densities for these distributions are given in table IV. The variate generation procedures are straight forward and can be found in Johnson (1985).

The critical values based on Foutz' asypmtotic normal distribution were used for F_n. Critical values for Hawkins' test statistic came from his recommended asymptotic values. The choice of critical values for the MAKS test statistic includes the recommended K-S values, referred to as MAKS1, and the empirical values from table II, referred to as MAKS2. The empirical values obviously have good test size and

TABLE IV

DISTRIBUTION DENSITIES

Plackett's Bivariate with Uniform Marginals

$$f(x_1,x_2) = \frac{\psi(\psi+1)(x_1+x_2-2x_1x_2)+1}{[\{1+(\psi-1)(x_1+x_2)\}^2-4\psi(\psi-1)x_1x_2]^{3/2}}$$

$\psi>0 \quad 0<x_1, \quad x_2<1$

Morgenstern's Bivariate with Uniform Marginals

$$f(x_1,x_2) = 1+\alpha(2x_1-1)(2x_2-1)$$

$0\leq x_1, \quad x_2\leq 1 \quad -1\leq\alpha\leq 1$

Khinchine's Theorem

$$f(x_1,x_2) = \frac{\alpha x_1x_2}{2A}\Phi(A) + \frac{1-\alpha x_1x_2}{2}[1-\Phi(A)]$$

$A = \max\{|x_1|,|x_2|\}$, $\quad\Phi$ is standard normal density

$-1\leq\alpha\leq 1$

Pearson Type II

$$f(\underline{x}) = \frac{\Gamma(p/2+m+1)}{\pi^{p/2}\Gamma(m+1)}|\Sigma|^{-1/2}[1-(\underline{x}-\underline{\mu})'\Sigma^{-1}(\underline{x}-\underline{\mu})]^m$$

$m>-1$

Pearson Type VII

$$f(\underline{x}) = \frac{\Gamma(m)}{\pi^{p/2}\Gamma(m-p/2)}|\Sigma|^{-1/2}[1+(\underline{x}-\underline{\mu})'\Sigma^{-1}(\underline{x}-\underline{\mu})]^{-m}$$

$m>1$

Johnson Translation System -- Lognormal Translation

$$f(x_1,x_2) = \frac{b_1 b_2}{2\pi\sigma_{12}(b_1x_1+a_1)(b_2x_2+a_2)(1-\rho^2)^{1/2}}$$

$$\times \exp -\frac{(y_1/\sigma_1)^2-2\rho y_1 y_2/(\sigma_1\sigma_2)+(y_2/\sigma_2)^2}{2(1-\rho^2)}$$

$x_i>-a_i/b_i, \quad y_i = \ln(b_i x_i + a_i)$

$a_i = \exp(\sigma_i{}^2/2)$

$b_i = [\exp(2\sigma_i{}^2)-\exp(\sigma_i{}^2)]^{1/2}$ for $i=1,2$

Johnson Translation System -- Sinh^{-1} Translation

$$f(x_1,x_2) = \frac{b_1 b_2[w_1+(1+w_1{}^2)^{1/2}][w_2+(1+w_2{}^2)^{1/2}]}{[1+w_1{}^2+w_1(1+w_1{}^2)^{1/2}][1+w_2{}^2+w_2(1+w_2{}^2)^{1/2}]g[\sinh^{-1}(w_1),\sinh^{-1}(w_2)]}$$

$w_i = b_i x_i + a_i, \quad a_i = \exp(\sigma_i{}^2/2)\sinh(\mu_i),$

$b_i = \{[\exp(\sigma_i{}^2)-1][\exp(\sigma_i{}^2)\cosh(2\mu_i)+1]\}^{1/2}$

$$g(y_1,y_2) = \frac{\exp\{-(y_1{}^2-2\rho y_1 y_2+y_2{}^2)/[2(1-\rho^2)]\}}{2\sigma_1\sigma_2(1-\rho^2)^{1/2}}$$

for $i=1,2,$

$\sinh^{-1}(x) = \ln[x+(1+x^2)^{1/2}]$

also exhibit 'super' power in detecting the slightest deviations from normality. The graphs in figures 3-9 and tables V and VI compare the power of the three tests with both choices of critical values for the MAKS test.

With the exception of the Johnson translation system (see tables V and VI), F_n is independent of location and scale changes. Therefore, the distribution location parameters are zero and scale parameters are one.

Table VII gives the percent rejection for the three tests for α=.01, .05, and .10 for the normal or near normal cases of the above distributions. All three tests are conservative with Hawkins being the most and Foutz the least conservative. Only the MAKS test with the empirical critical values is capable of detecting the slight departures from normality.

Figures 3-9 are plots of the percent rejection for α=.05 versus the key parameters of each distribution. The normal or near-normal cases are designated by an asterisk above the set of symbols. The solid curves are spline fits of the displayed points and serve only as a visual guide. Following the graphs (figures 3-8) are 3-dimensional density plots and projected contour plots of the the distributions (from Johnson, 1985). In almost all cases, F_n is not as powerful as the other two tests. Typically F_n has about half the power of MAKS and Hawkins.

TABLE V

PERCENT REJECTION FOR JOHNSON TRANSLATION SYSTEM AT 5%

$SINH^{-1}$ TRANSLATION

μ_1	μ_2	σ_1	σ_2	ρ	Foutz	Hawkins	MAKS1	MAKS2
0	0	1.0	1.0	-.5	48.6	81.8	81.9	92.2
0	0	1.0	1.0	0.0	43.8	79.4	79.3	90.5
0	0	1.0	1.0	0.5	48.9	81.4	82.2	93.1
0	0	1.0	1.0	0.8	50.5	81.8	83.2	92.1
0	0	0.1	0.1	-.5	[illegible].5	0.4	0.8	5.2*
0	0	0.1	0.1	0.0	2.8	0.4	1.3	4.5*
0	0	0.1	0.1	0.5	5.0	0.3	1.3	5.2*
0	0	0.1	0.1	0.8	3.0	0.7	1.2	5.4*
0	2	1.0	1.0	-.5	59.0	88.1	92.0	94.8
0	2	1.0	1.0	0.0	54.7	85.1	88.7	94.3
0	2	1.0	1.0	0.5	66.4	89.8	92.7	96.9
0	2	1.0	1.0	0.8	59.8	86.1	89.4	96.8
0	0	0.1	1.0	-.5	11.1	29.4	30.7	51.8
0	0	0.1	1.0	0.0	11.4	26.5	26.7	50.1
0	0	0.1	1.0	0.5	11.9	28.9	31.1	51.9
0	0	0.1	1.0	0.8	12.1	36.2	37.3	57.4
0	2	0.1	0.5	-.5	4.9	6.7	8.8	24.1
0	2	0.1	0.5	0.0	5.2	5.2	8.4	21.4
0	2	0.1	0.5	0.5	6.8	6.0	10.0	25.6
0	2	0.1	0.5	0.8	7.1	10.9	14.4	32.9
0	2	0.1	1.0	-.5	16.7	43.1	51.5	72.2
0	2	0.1	1.0	0.0	16.4	42.5	50.7	71.3
0	2	0.1	1.0	0.5	17.8	46.7	55.5	73.9
0	2	0.1	1.0	0.8	21.4	52.7	59.5	76.8
0	2	1.0	0.5	-.5	25.5	50.8	50.8	70.4
0	2	1.0	0.5	0.0	23.8	50.7	54.2	74.3
0	2	1.0	0.5	0.5	25.6	52.2	54.9	75.5
0	2	1.0	0.5	0.8	30.8	61.4	64.7	79.7
2	2	0.5	1.0	-.5	36.4	66.6	73.1	86.1
2	2	0.5	1.0	0.0	37.9	68.1	76.4	88.4
2	2	0.5	1.0	0.5	42.6	73.1	81.4	91.5
2	2	0.5	1.0	0.8	46.4	77.4	83.2	91.8
2	2	1.0	1.0	-.5	87.7	91.9	95.3	98.3
2	2	1.0	1.0	0.0	75.6	91.6	95.6	98.5
2	2	1.0	1.0	0.5	81.6	96.4	98.6	99.4
2	2	1.0	1.0	0.8	87.7	97.3	96.2	99.3
2	2	0.5	0.5	-.5	14.2	24.8	32.1	52.1
2	2	0.5	0.5	0.0	18.0	26.3	33.6	54.6
2	2	0.5	0.5	0.5	17.5	29.6	38.8	60.1
2	2	0.5	0.5	0.8	21.3	37.0	46.8	65.7

TABLE VI

PERCENT REJECTION FOR JOHNSON TRANSLATION SYSTEM AT 5%

LOGNORMAL TRANSLATION

μ_1	μ_2	σ_1	σ_2	ρ	Foutz	Hawkins	MAKS1	MAKS2
0	0	.05	.05	0.0	4.0	0.2	0.1	5.5*
0	0	.05	.05	-.5	3.6	0.6	0.1	5.8*
0	0	.05	.05	0.5	3.6	0.7	1.3	5.3*
0	0	.05	.05	0.8	3.2	0.3	0.9	4.8*
0	0	0.5	1.0	-.5	19.6	44.4	46.0	66.7
0	0	0.5	1.0	0.0	20.2	42.5	44.6	66.5
0	0	0.5	1.0	0.5	18.2	46.0	46.7	67.3
0	0	0.5	1.0	0.8	19.8	49.5	51.7	71.3
0	0	.05	1.0	-.5	11.8	27.2	28.7	50.5
0	0	.05	1.0	0.0	9.7	23.8	24.0	47.5
0	0	.05	1.0	0.5	12.4	28.9	31.0	51.6
0	0	.05	1.0	0.8	14.0	33.1	35.6	56.6
0	0	.05	0.5	-.5	3.1	1.5	2.6	11.1
0	0	.05	0.5	0.0	5.1	1.8	3.0	12.0
0	0	.05	0.5	0.5	3.7	2.0	3.3	12.2
0	0	.05	0.5	0.8	3.0	0.4	0.9	13.5
0	0	0.5	0.5	-.5	6.9	8.2	10.9	25.8
0	0	0.5	0.5	0.0	5.3	5.4	8.9	25.4
0	0	0.5	0.5	0.5	5.6	6.9	10.1	25.8
0	0	0.5	0.5	0.8	6.4	7.8	10.6	26.1
0	0	1.0	1.0	-.5	46.4	83.2	83.1	93.2
0	0	1.0	1.0	0.0	43.0	79.4	79.5	90.7
0	0	1.0	1.0	0.5	44.9	81.5	81.1	91.5
0	0	1.0	1.0	0.8	46.6	82.4	81.6	92.3

*denotes near normal cases

TABLE VII

PERCENT REJECTION OF THE NORMAL HYPOTHESIS FOR

THE NORMAL OR NEAR-NORMAL CASES, 1000 SAMPLES OF SIZE 50

Distribution	Level	Foutz	Hawkins	MAKS1	MAKS2
Normal	10%	6.70	0.70	2.50	8.80
	5%	2.90	0.30	0.80	3.50
	1%	0.60	0.00	0.10	1.00
Khintchine*	10%	9.50	8.40	14.3	27.8
	5%	4.70	4.20	6.40	18.6
	1%	1.10	0.50	0.80	6.90
$Sinh^{-1}$	10%	6.20	1.20	2.90	8.00
	5%	2.80	0.40	1.30	4.50
	1%	0.80	0.00	0.00	1.40
Plackett	10%	6.00	1.70	2.90	9.20
	5%	2.40	0.40	1.10	4.40
	1%	0.60	0.00	0.00	1.20
Morgenstern	10%	7.50	1.00	2.70	8.20
	5%	3.70	0.20	0.20	4.00
	1%	0.40	0.00	0.00	0.50
Lognormal	10%	7.00	0.80	3.30	10.1
	5%	4.00	0.20	0.10	5.80
	1%	0.80	0.00	0.00	0.50
Pearson VII	10%	7.50	4.20	7.70	19.6
	5%	3.80	2.20	3.30	11.8
	1%	1.00	0.10	0.70	3.90
Pearson II	10%	7.00	2.40	4.90	13.7
	5%	2.60	0.70	1.60	8.00
	1%	0.30	0.00	1.60	1.80

*This is distribution does not have a very near normal case.

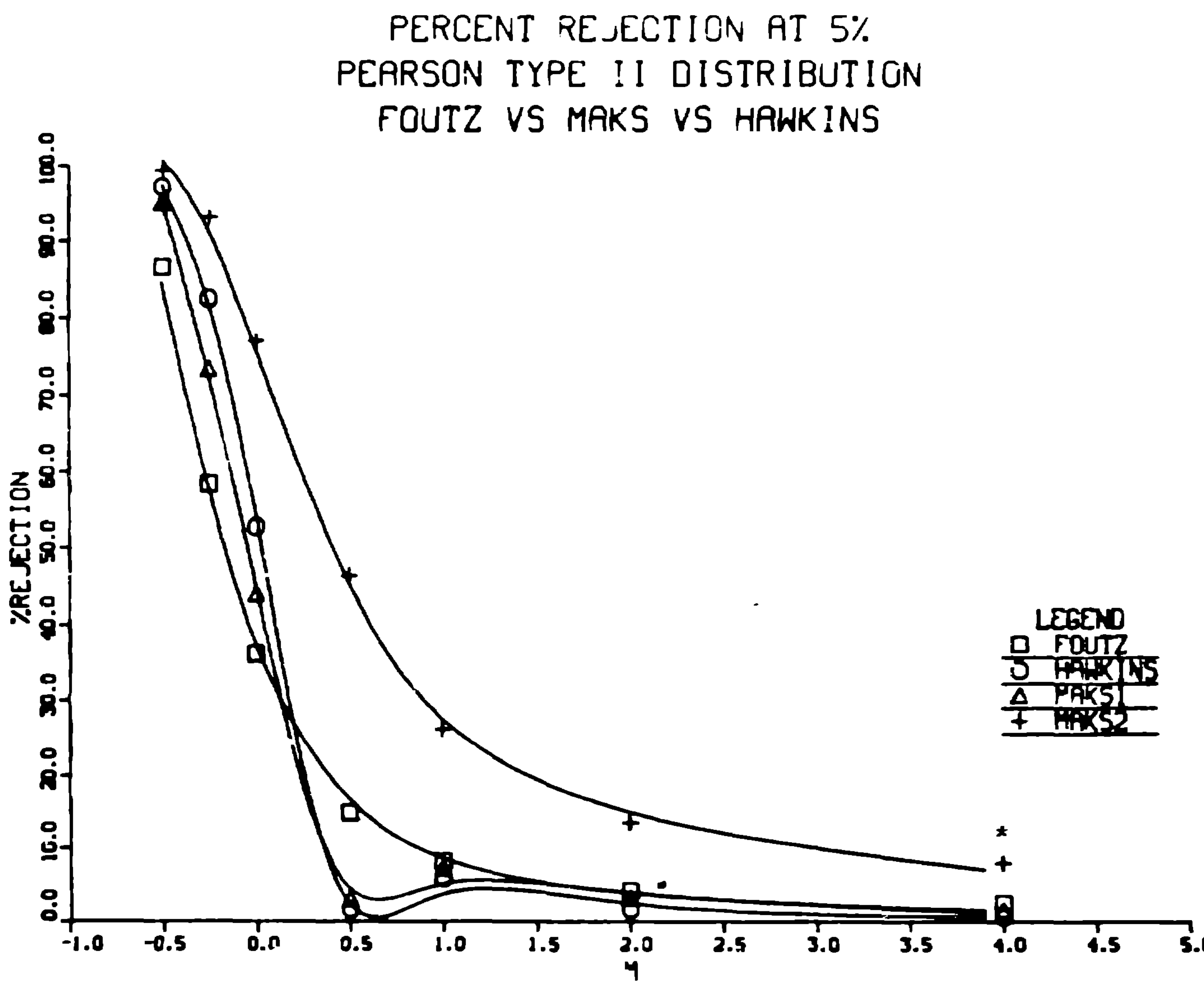

PERCENT REJECTION AT 5%
PEARSON TYPE II DISTRIBUTION
FOUTZ VS MAKS VS HAWKINS
%REJECTION
LEGEND
FOUTZ
HAWKINS
MAKS1
MAKS2

Figure 3A

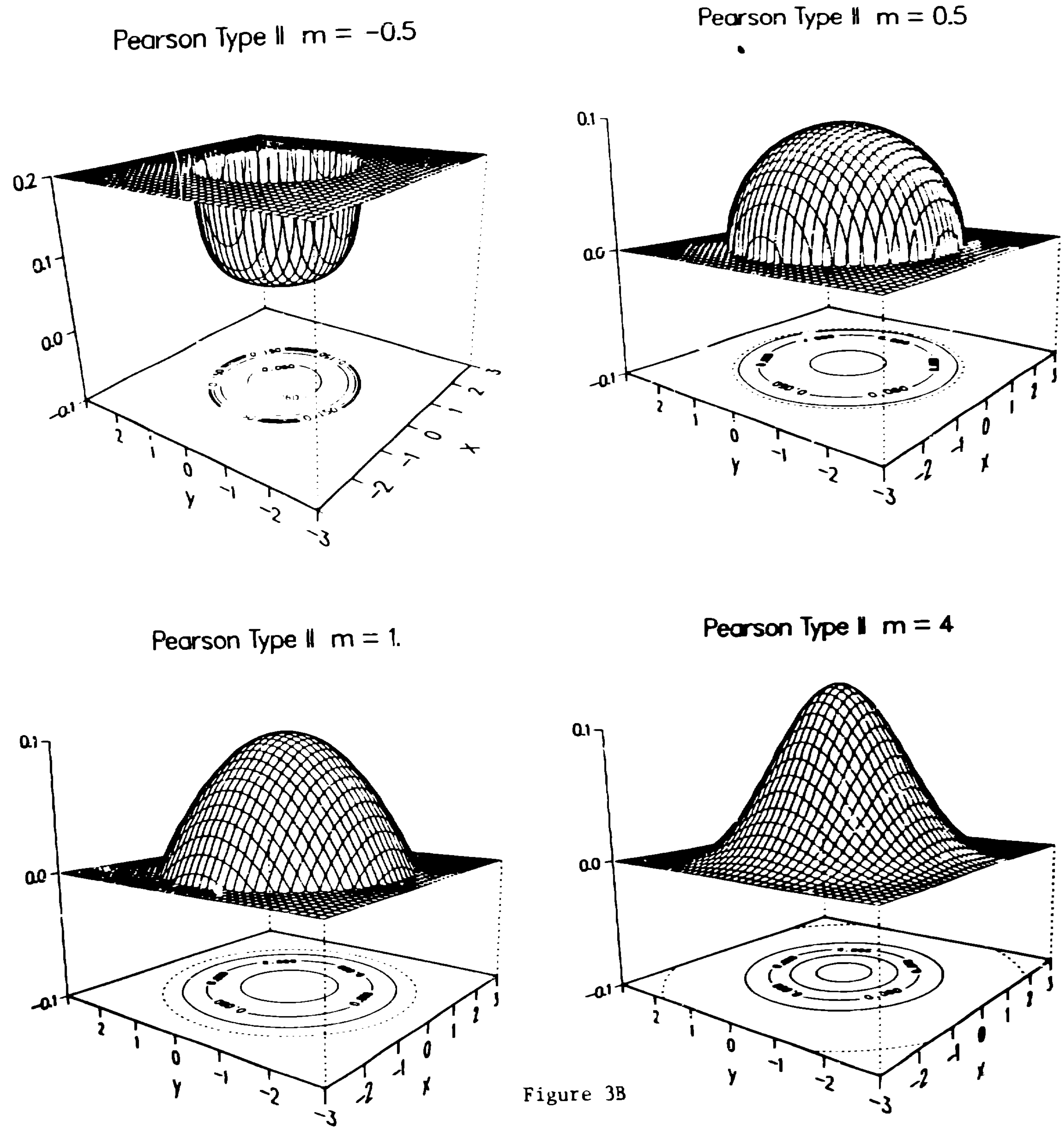

Pearson Type II m = −0.5
Pearson Type II m = 0.5
Pearson Type II m = 1.
Pearson Type II m = 4
y
x

Figure 3B

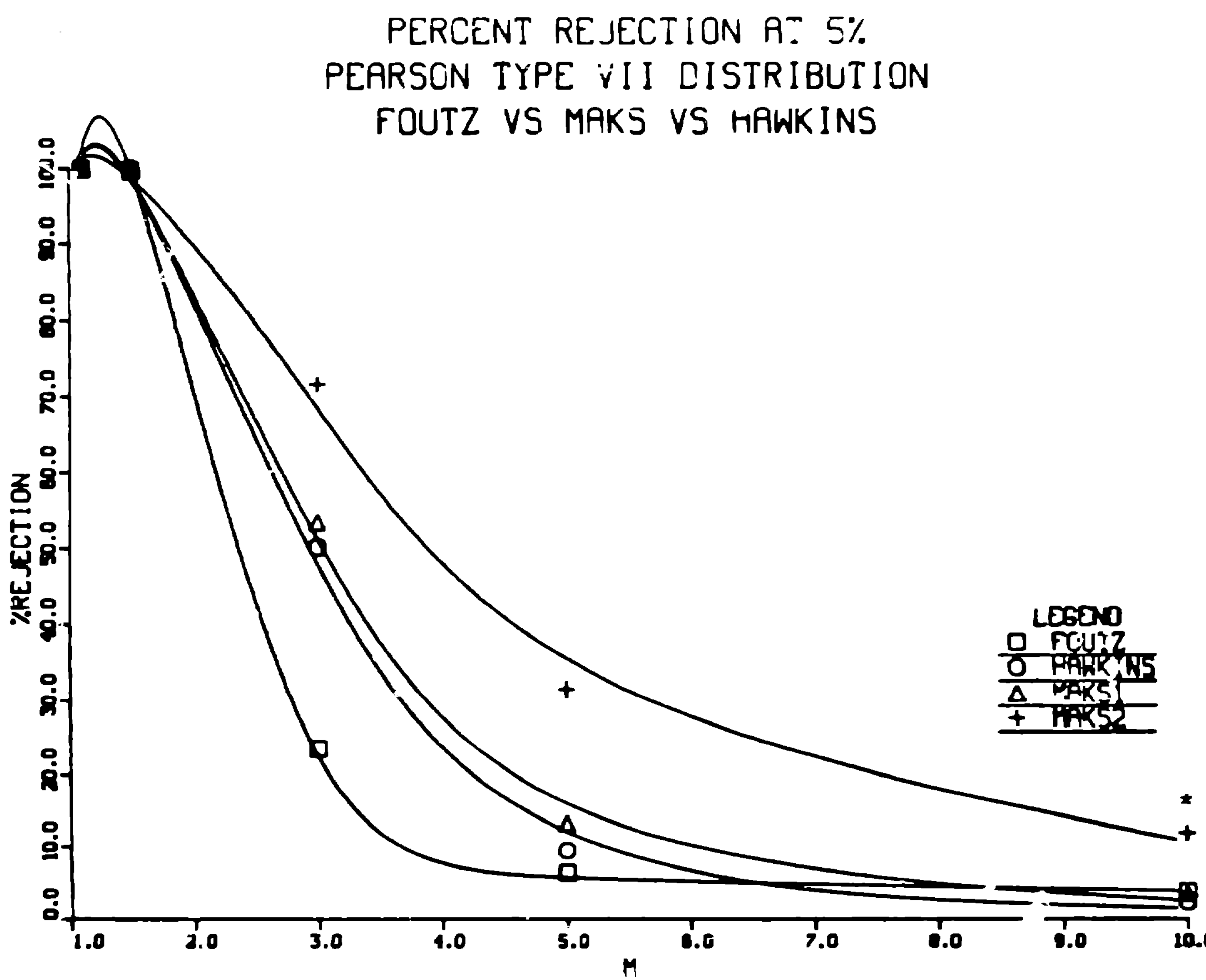
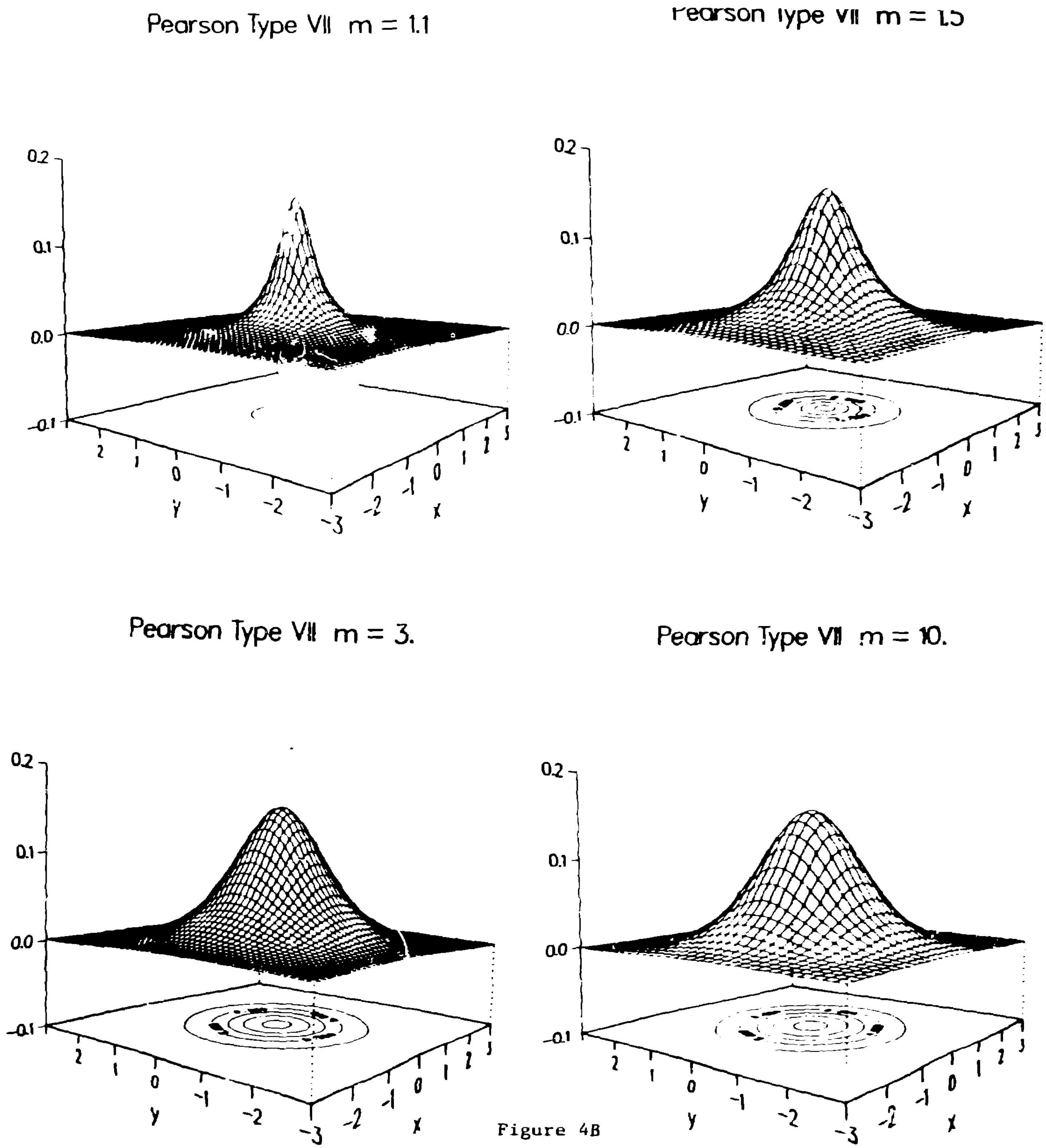

PERCENT REJECTION AT 5%
PEARSON TYPE VII DISTRIBUTION
FOUTZ VS MAKS VS HAWKINS
%REJECTION
0.0
10.0
20.0
30.0
40.0
50.0
60.0
70.0
80.0
90.0
100.0
1.0
2.0
3.0
4.0
5.0
6.0
7.0
8.0
9.0
10.0
M
LEGEND
FOUTZ
HAWKINS
MAKS1
MAKS2

Figure 4A



Pearson Type VII m = 1.1
Pearson Type VII m = 1.5
Pearson Type VII m = 3.
Pearson Type VII m = 10.
0.2
0.1
0.0
-0.1
y
x

Figure 4B

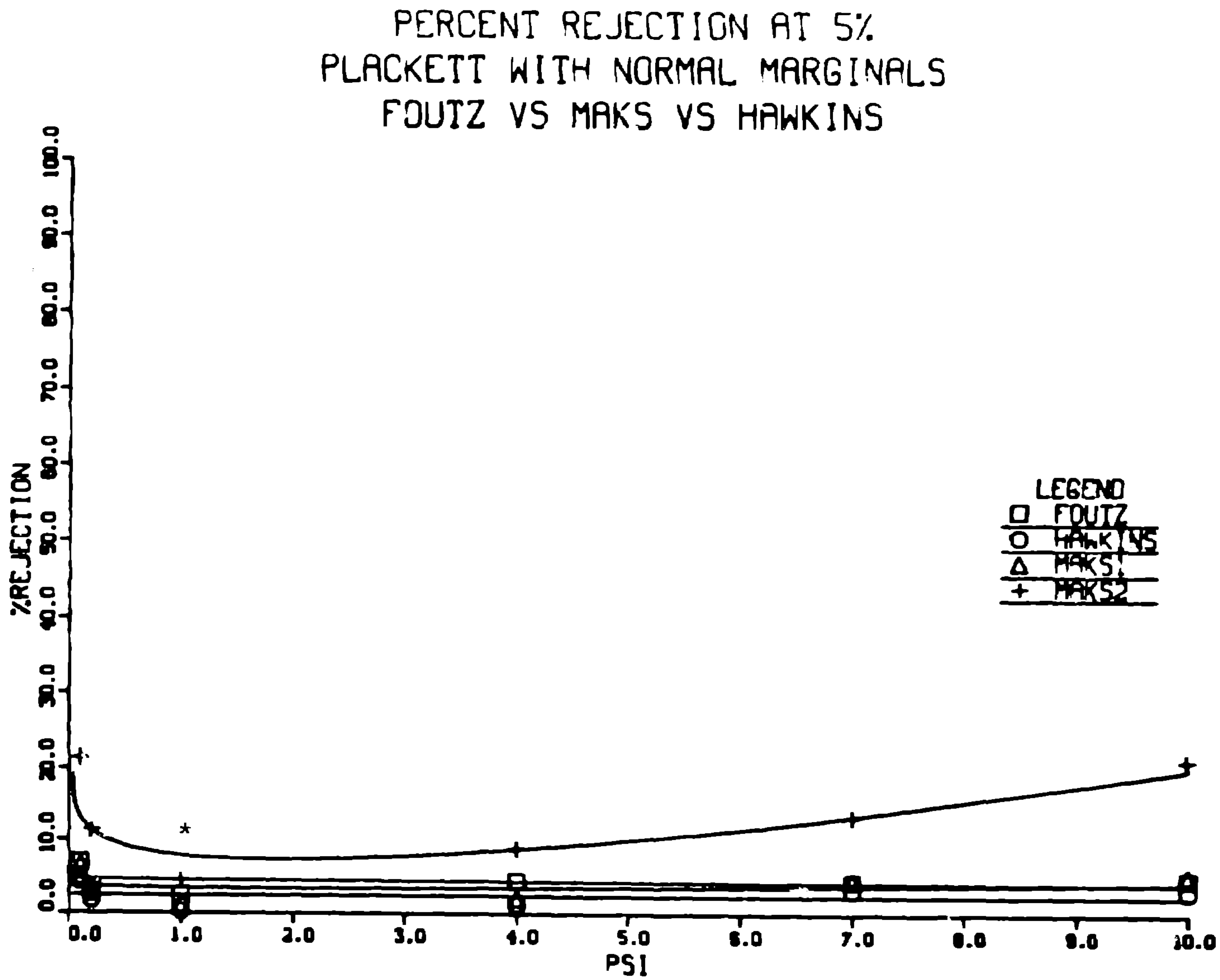

PERCENT REJECTION AT 5%
PLACKETT WITH NORMAL MARGINALS
FOUTZ VS MAKS VS HAWKINS
%REJECTION
PSI
LEGEND
FOUTZ
HAWKINS
MAKS1
MAKS2

Figure 5A

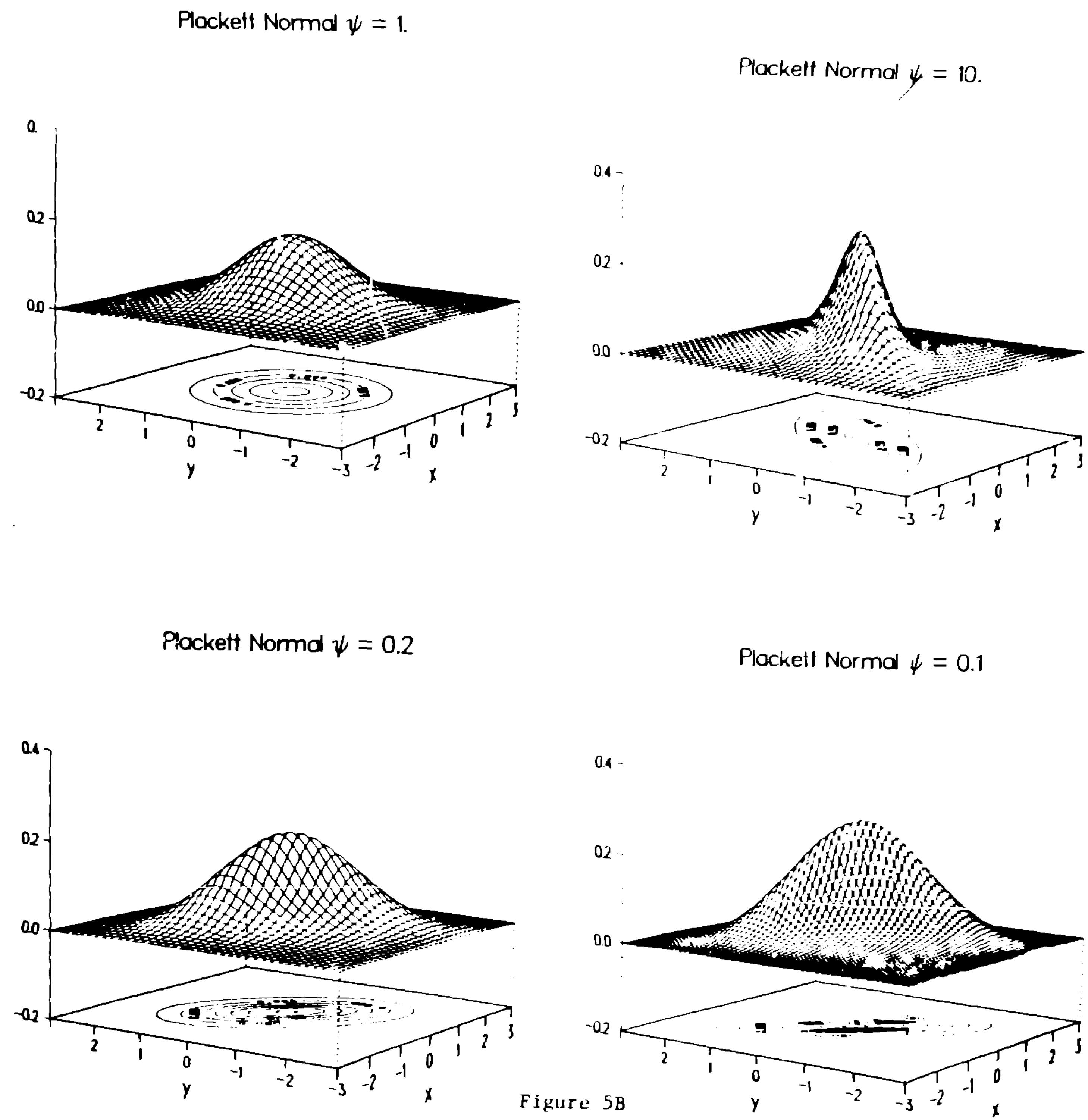

Plackett Normal $\psi = 1.$
Plackett Normal $\psi = 10.$
Plackett Normal $\psi = 0.2$
Plackett Normal $\psi = 0.1$
0.4
0.2
0.0
-0.2
y
x

Figure 5B

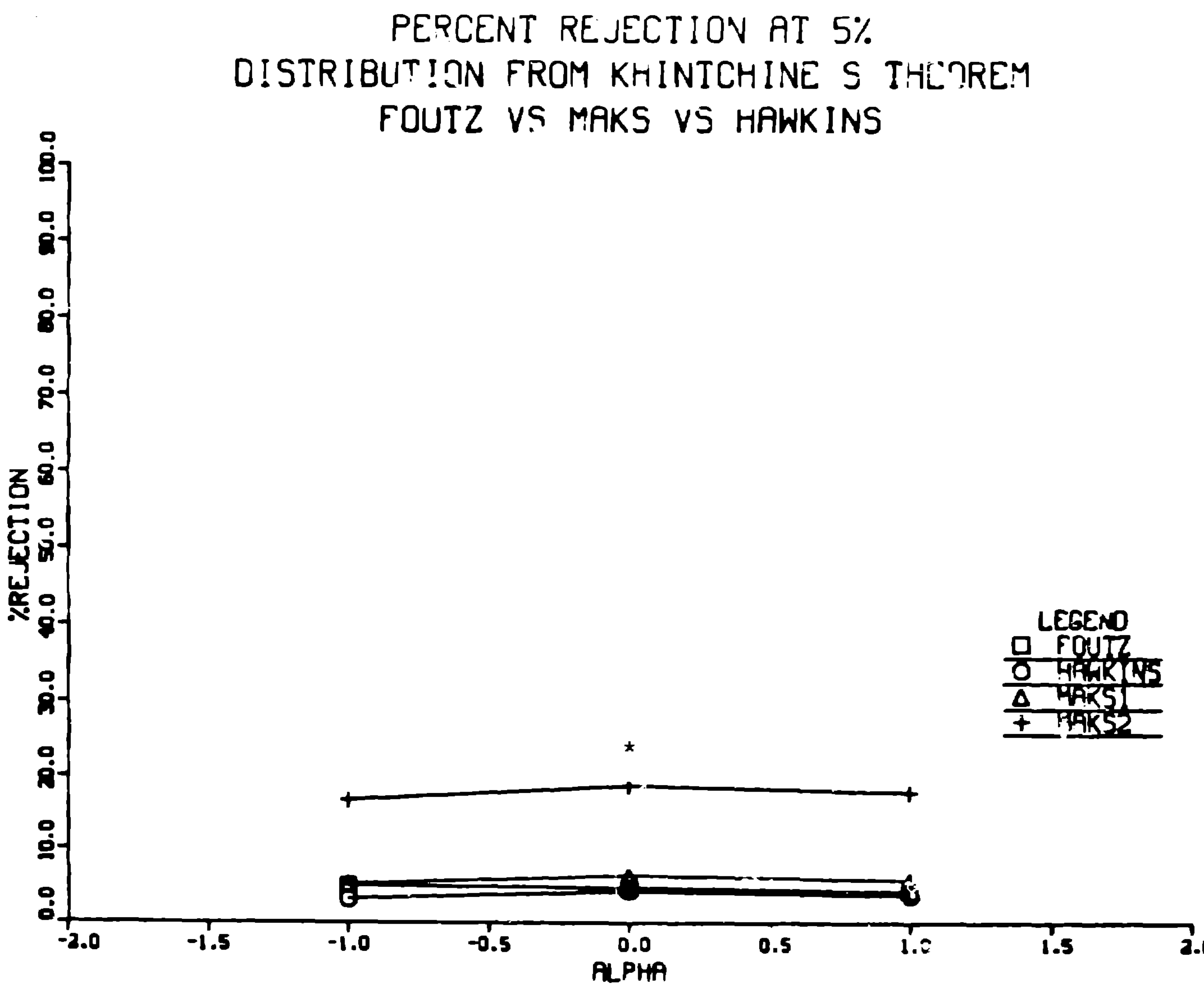
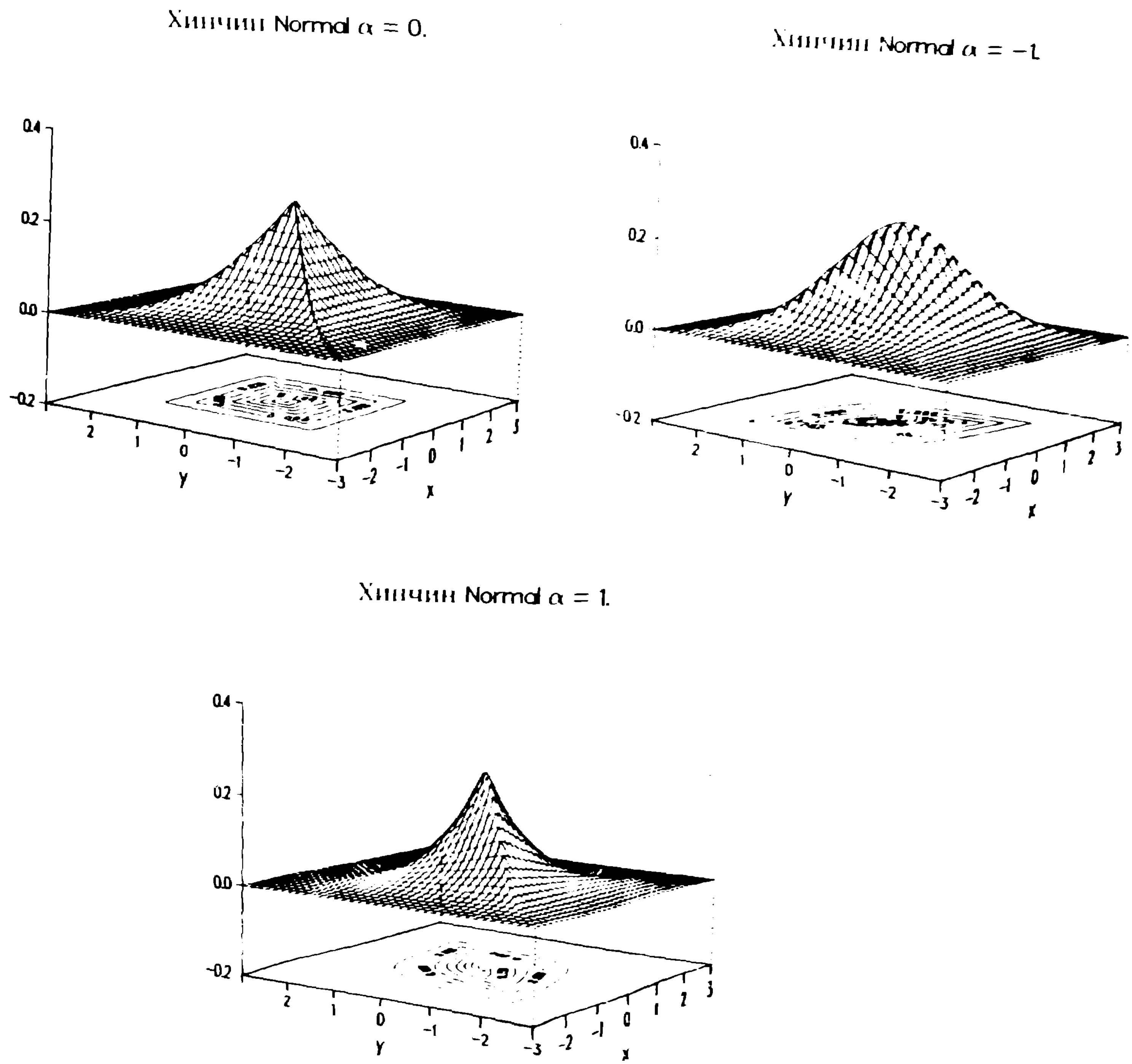

PERCENT REJECTION AT 5%
DISTRIBUTION FROM KHINTCHINE S THEOREM
FOUTZ VS MAKS VS HAWKINS
%REJECTION
0.0
10.0
20.0
30.0
40.0
50.0
60.0
70.0
80.0
90.0
100.0
-2.0
-1.5
-1.0
-0.5
0.0
0.5
1.0
1.5
2.0
ALPHA
LEGEND
FOUTZ
HAWKINS
MAKS1
MAKS2

Figure 6A



Хинчин Normal α = 0.
Хинчин Normal α = −1.
Хинчин Normal α = 1.
0.4
0.2
0.0
−0.2
y
x

Figure 6B

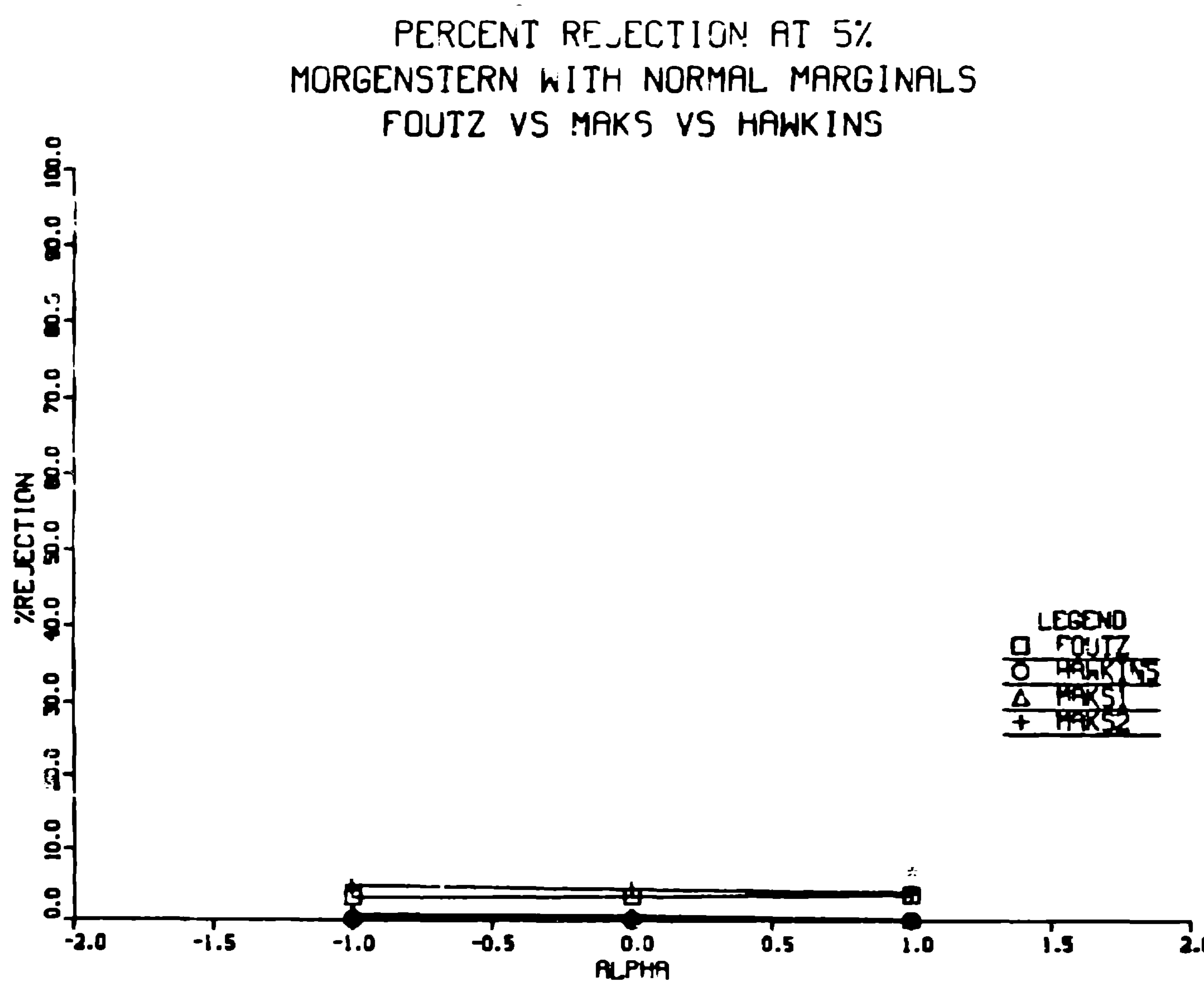

PERCENT REJECTION AT 5%
MORGENSTERN WITH NORMAL MARGINALS
FOUTZ VS MAKS VS HAWKINS
%REJECTION
ALPHA
LEGEND
FOUTZ
HAWKINS
MAKS1
MAKS2

Figure 7A

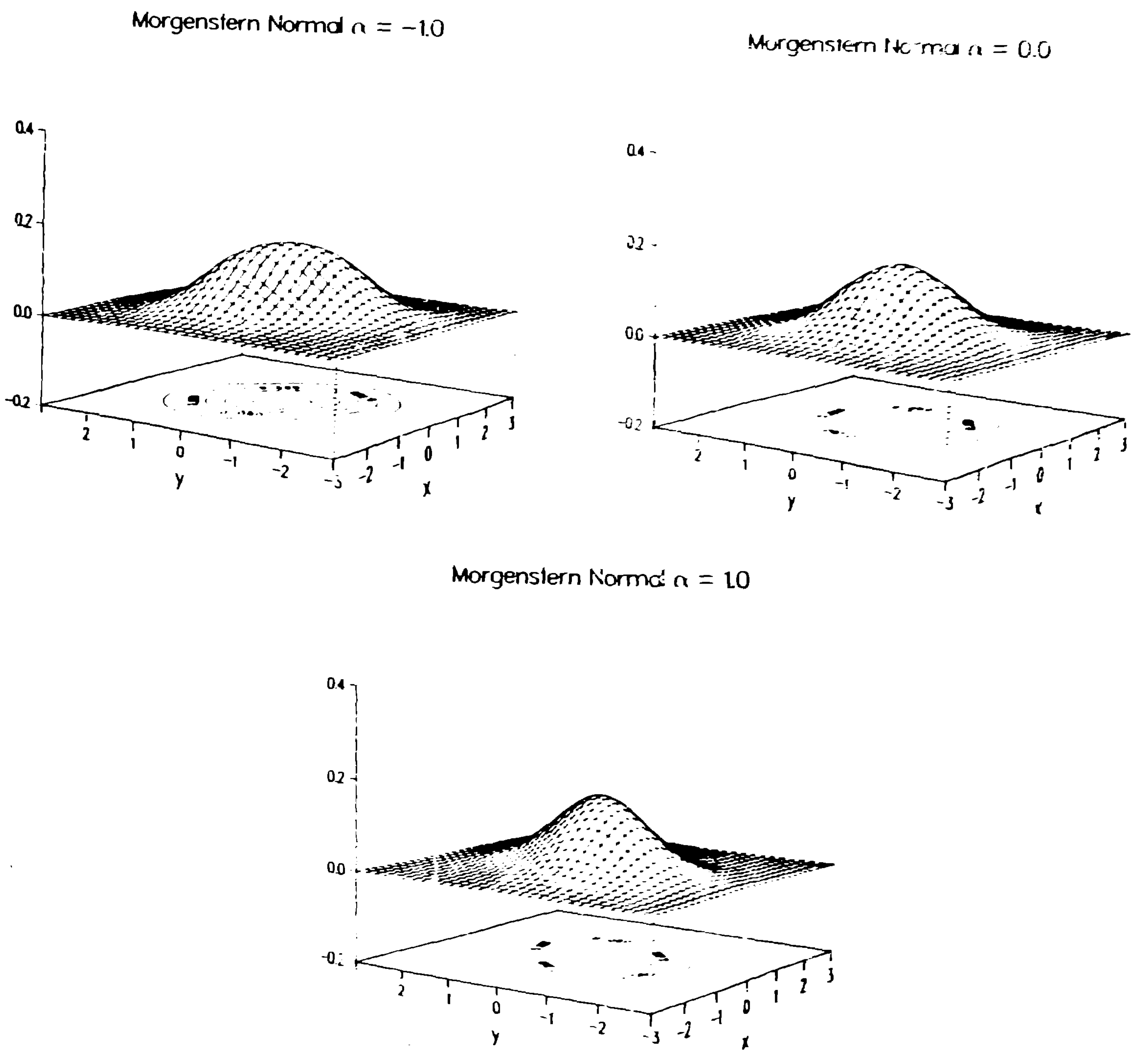

Morgenstern Normal α = −1.0
0.4
0.2
0.0
-0.2
2
1
0
-1
-2
-3
-2
-1
0
1
2
3
y
x
Morgenstern Normal α = 0.0
0.4
0.2
0.0
-0.2
y
x
Morgenstern Normal α = 1.0
0.4
0.2
0.0
y
x

Figure 7B

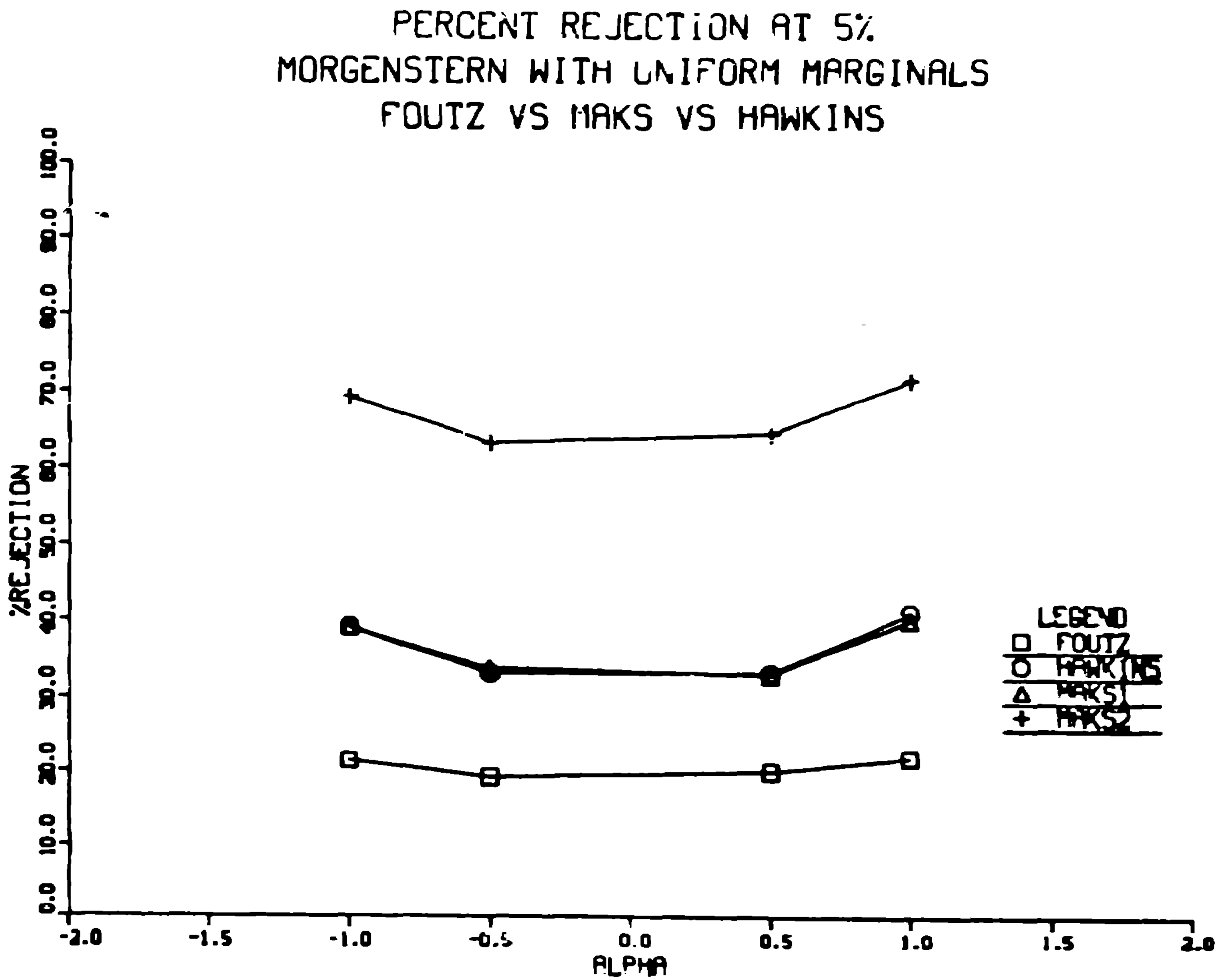

PERCENT REJECTION AT 5%
MORGENSTERN WITH UNIFORM MARGINALS
FOUTZ VS MAKS VS HAWKINS
%REJECTION
ALPHA
LEGEND
FOUTZ
HAWKINS
MAKS1
MAKS2
-2.0
-1.5
-1.0
-0.5
0.0
0.5
1.0
1.5
2.0

Figure 8A

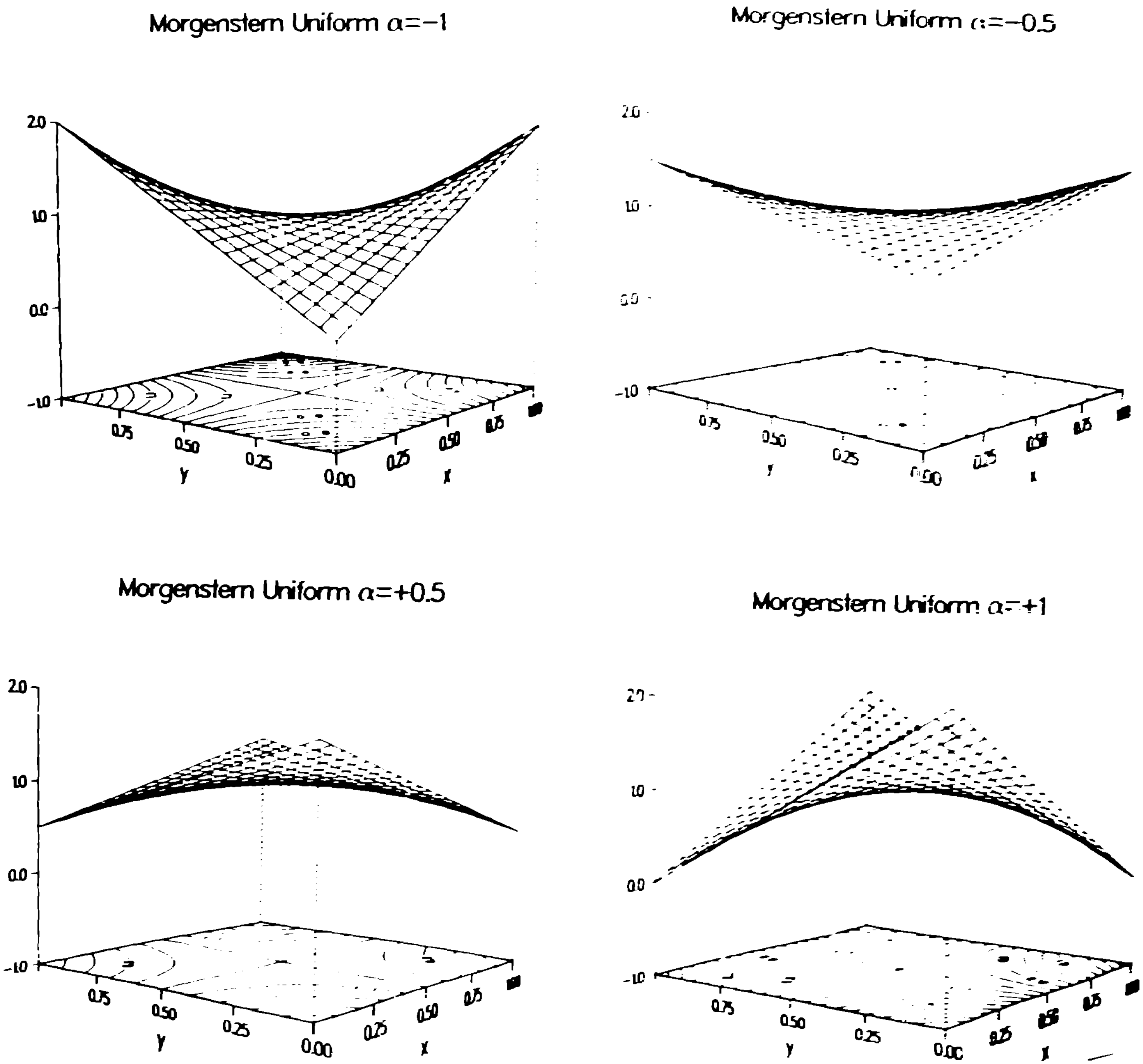

Morgenstern Uniform α=−1
Morgenstern Uniform α=−0.5
Morgenstern Uniform α=+0.5
Morgenstern Uniform α=+1
2.0
1.0
0.0
−1.0
0.75
0.50
0.25
0.00
y
x

Figure 8B

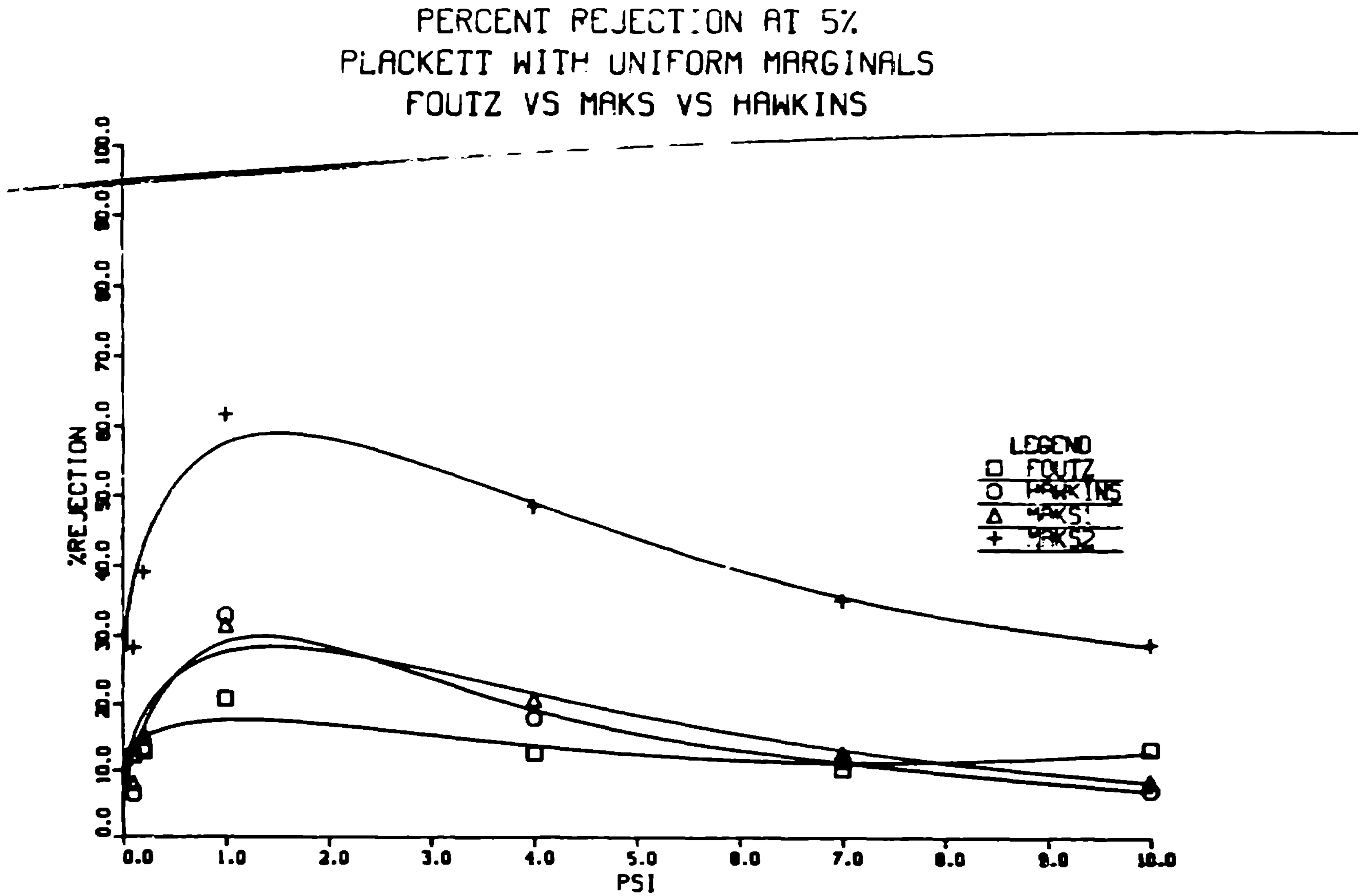

PERCENT REJECTION AT 5%
PLACKETT WITH UNIFORM MARGINALS
FOUTZ VS MAKS VS HAWKINS
%REJECTION
PSI
0.0
1.0
2.0
3.0
4.0
5.0
7.0
10.0
LEGEND
FOUTZ
HAWKINS
MAKS1
MAKS2

Figure 9

VII. CONCLUSIONS

In general the power of F_n is relatively poor compared to multivariate normal tests developed by Malkovich-Afifi and Hawkins in both the univariate and bivariate cases. However, empirical studies of F_n show that it reasonably follows its asymptotic distribution and is indeed independent of p. The Hawkins test and MAKS1 test (using K-S critical values) have comparable performance and do well under the conditions tested. The MAKS2 test (using empirical critical values) is consistently the most powerful of all these multivariate normality tests against all alternatives.

In spite of the rather discouraging performance of Foutz' test, we feel that this study accomplished the following:

1) Objective appraisal of the Foutz test and resolution of the discrepancies in Foutz (1980) with Franke and Jayachandran (1983) with regard to power against a broad family of univariate distributions.

2) Adaptation of the Foutz procedure to handle an estimated mean vector and covariance matrix. The distribution-free aspect of the Foutz test was observed in this extended study.

3) The providing of a set of candidate alternative distributions for use in studies such as these.

REFERENCES

Anderson, T. W., Some Nonparametric Procedures Based on Statistically Equivalent Blocks, Procedings of the International Symposium on Multivariate Analysis; pp. 5-27, New York, Academic Press, 1966.

Foutz, R. V., A Test for Goodness-of-Fit Based on an Empirical Probabilistic Measure, Annals of Statistics; vol. 8, pp. 989-1001, 1980.

Franke, R. and T. Jayachandran, An Empirical Investigation of the Properties of a New Goodness-of-Fit Test, Journal of Statistical Computation and Simulation; vol. 16, pp. 175-187, 1983.

Hawkins, D. M., A New Test for Multivariate Normality and Homoscedasticity, Technometrics; vol. 23, pp. 105-110, 1981.

Johnson, M. E., Multivariate Statistical Generation; to be published, 1985 (currently Los Alamos report LA-UR 84-495).

Johnson, M. E., G. L. Tietjen, and R. J. Beckman, A New Family of Probability Distributions with Applications to Monte Carlo Studies, Journal of the American Statistical Association; vol. 75, pp. 276-279, 1980.

Malkovich, J. F. and A. A. Afifi, On Tests for Multivariate Normality, Journal of the American Statistical Association; vol. 68, pp. 176-179, 1973.